Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Linear Algebra and its Applications 388 (2004) 289–294

LINEAR ALGEBRA AND ITS APPLICATIONS

www.elsevier.com/locate/laa

A note on a matrix inequality for generalized means

Alexander Kukush [a], Hans Schneeweiss [b,*]

[a] *Department of Mathematics, Kiev Taras Shevchenko University, Ukraine*
[b] *Department of Statistics, University of Munich, Akademiestr. 1, Munich 80799, Germany*

Received 18 February 2003; accepted 12 January 2004

Submitted by H.J. Werner

Abstract

In Statistics, generalized means of positive random variables are often considered. As is well-known, the generalized mean of order α, $\{E(v^\alpha)\}^{\frac{1}{\alpha}}$, is smaller (greater) than the ordinary mean if $\alpha < 1$ ($\alpha > 1$). This result can be generalized to a corresponding inequality involving matrix random variables of a specific type. In the special case when $\alpha = -1$, we have a matrix inequality that has applications in various fields of Statistics. Two such applications are presented.

Keywords: Generalized means; Jensen's inequality; Loewner order; Measurement errors; Poisson model; Polynomial model

1. Introduction

Consider a positive random variable v. Its generalized mean of order α, $\alpha \neq 0$, is given by

$$\mu_\alpha = \{E(v^\alpha)\}^{\frac{1}{\alpha}}.$$

For $\alpha = 1$ we have the ordinary mean, for $\alpha = 2$ the quadratic mean, and for $\alpha = -1$ the harmonic mean. The definition of μ_α can be extended to the case $\alpha = 0$:

* Corresponding author. Fax: +49-89-2180-5044.

E-mail addresses: schneew@stat.uni-muenchen.de, schneew@stat.uni-muenchen.edu (H. Schneeweiss).

0024-3795/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.laa.2004.01.004

$$\mu_0 = \exp(E(\ln v)),$$

which is the geometric mean of v.

Suppose v is not restricted to a single point, we then have the well-known inequality,

$$\mu_\alpha < \mu < \mu_\beta,$$

whenever $\alpha < 1 < \beta$, assuming that $\mu := Ev = \mu_1$ is finite, e.g. [1, Section 2.6].

This inequality can be generalized to the case where v is multiplied by a matrix of the form ww', w being a random vector with $Eww' = I$.

2. The matrix inequality

The following lemma generalizes the concept of a generalized mean of order α.

Lemma. *Suppose that x, y are random variables with $x \geqslant 0$ and $y > 0$ a.s., $Ex = 1$, $E(xy) < \infty$, and for any $d > 0$, $P(y \neq d, x \neq 0) > 0$. Then, for any $\alpha < 1$ and $\beta > 1$,*

$$\{E(xy^\alpha)\}^{\frac{1}{\alpha}} < E(xy) < \{E(xy^\beta)\}^{\frac{1}{\beta}}, \tag{1}$$

where, by definition,

$$\{E(xy^\alpha)\}^{\frac{1}{\alpha}}\Big|_{\alpha=0} := \mathrm{e}^{E(x \ln y)}.$$

Proof. Let $F(x, y)$ be the joint d.f. of x and y, then for each $c \neq 0$,

$$E(xy^c) = \int_{\mathbb{R}} y^c \, \mathrm{d}G(y), \tag{2}$$

with

$$G(y) = \int_{\mathbb{R}} x \, \mathrm{d}F(x, y), \quad y \in \mathbb{R}.$$

The function $G(y)$ is a probability d.f. since,

$$G(+\infty) = Ex = 1.$$

Denote the integral on the right hand side of (2) by $E_x y^c$. Note that the distribution $G(y)$ is not concentrated in a single point $y = d$ because this would imply $P(y \neq d, x \neq 0) = 0$ contrary to the assumption. For $\alpha \neq 0$ the inequality (1) can be written in the form

$$(E_x y^\alpha)^{\frac{1}{\alpha}} < E_x y < (E_x y^\beta)^{\frac{1}{\beta}}$$

and follows from Jensen's inequality, because $E_x y = E(xy) < \infty$. For $\alpha = 0$ we have to show that

$$e^{E_x \ln y} < E_x y, \tag{3}$$

where

$$E_x \ln y := \int_{\mathbb{R}} \ln y \, \mathrm{d}G(y).$$

But this also follows from Jensen's inequality, which completes the proof. $\square$

Proposition. *Let v be a positive random variable with a distribution which has no atoms and w be a random (column) vector in $\mathbb{R}^m$, with $Eww' = I_m$. Let α and β be real numbers as previously stipulated and assume $E(\|w\|^2 v) < \infty$. Then, for $\alpha \neq 0$,*

$$\operatorname{tr}\{E(ww'v^\alpha)\}^{\frac{1}{\alpha}} < \operatorname{tr} E(ww'v) < \operatorname{tr}\{E(ww'v^\beta)\}^{\frac{1}{\beta}}. \tag{4}$$

This inequality holds true also for $\alpha = 0$, where by definition

$$\{E(ww'v^\alpha)\}^{\frac{1}{\alpha}}\Big|_{\alpha=0} := e^{E(ww' \ln v)}.$$

Proof. Assume $\alpha \neq 0$ and <1. First note that, by the lemma, $E(\|w\|^2 v^\alpha) < \infty$ (set $\frac{1}{m}\|w\|^2 = x$ and $v = y$). Now, let (λ, φ) be an eigenvalue and normalized eigenvector pair of $E(ww'v^\alpha)$. Then

$$\begin{aligned}\varphi'\{E(ww'v^\alpha)\}^{\frac{1}{\alpha}}\varphi &= \lambda^{\frac{1}{\alpha}} = \{\varphi' E(ww'v^\alpha)\varphi\}^{\frac{1}{\alpha}} \\ &= \{E(\varphi' w)^2 v^\alpha\}^{\frac{1}{\alpha}} = \{E(xv^\alpha)\}^{\frac{1}{\alpha}},\end{aligned}$$

where $x = (\varphi' w)^2$, and $Ex = \varphi'(Eww')\varphi = \varphi'\varphi = 1$. We can now show that for each $d > 0$, $P(v \neq d, x \neq 0) > 0$. For suppose $P(v \neq d$ and $x \neq 0) = 0$ or equivalently $P(v = d$ or $x = 0) = 1$, then $P(v = d) > 0$ because $P(v = d) = 0$ would imply $P(x = 0) = 1$, which is impossible since $Ex = 1$. But $P(v = d) > 0$ contradicts the assumption that the distribution of v has no atoms. From the lemma we have,

$$\{E(xv^\alpha)\}^{\frac{1}{\alpha}} < E(xv) = \varphi' E(ww'v)\varphi$$

and, therefore,

$$\varphi'\{E(ww'v^\alpha)\}^{\frac{1}{\alpha}}\varphi < \varphi' E(ww'v)\varphi.$$

Summing over all φ belonging to an eigenbasis of $E(ww'v^\alpha)$, we obtain the left hand side of (4). The right hand side is established in a similar way.

For the case $\alpha = 0$ we start with (λ, φ) being an eigenvalue and eigenvector pair of $E(ww' \ln v)$. By arguments analogous to the above, we find that,

$$\varphi' e^{E(ww' \ln v)}\varphi < \varphi' E(ww'v)\varphi,$$

from which again the left hand side of (4) follows for $\alpha = 0$. $\square$

For an earlier version of the proof see [2].

Remark 1. If the distribution of v has atoms, then (4) holds with nonstrict inequality.

Remark 2. For $\alpha = -1$, we have the stronger proposition,

$$\{E(ww'v^{-1})\}^{-1} \leqslant E(ww'v)$$

in the Loewner sense. For a proof see [4].

3. Statistical applications

We have two applications in the theory of measurement error models of the proposition and its stronger version for $\alpha = -1$, see Remark 2.

3.1. Poisson model

Shklyar and Schneeweiss [4] consider the log-linear Poisson model. It is given by a Poisson distributed random variable y with mean parameter λ, where $\log \lambda$ is a linear function of a random vector ξ: $\log \lambda = \beta_0 + \beta_1' \xi$ with an unknown parameter vector $\beta = (\beta_0, \beta_1')'$. Assume that $\xi \sim N(\mu_\xi, \Sigma_\xi)$ with known mean vector μ_ξ and covariance matrix Σ_ξ. Assume further that ξ is unobservable. Instead a random vector x is observed, which is related to ξ by the equation $x = \xi + \delta$, $\delta \sim N(0, \Sigma_\delta)$, δ being the vector of measurement errors with known covariance matrix Σ_δ. δ is assumed to be independent of ξ and y.

In this model, it is possible to evaluate the conditional mean and variance of y given x as functions of x and β:

$$E(y \mid x) = m(x, \beta)$$

$$V(y \mid x) = v(x, \beta).$$

They can by used to construct unbiased estimating functions $\psi(y, x, \beta)$ with the property that $E[\psi(y, x, \beta)] = 0$. In particular the following two estimating functions will be considered:

$$\psi_1(y, x, \beta) = [y - m(x, \beta)]v^{-1}(x, \beta)m_\beta(x, \beta),$$

$$\psi_2(y, x, \beta) = [y - m(x, \beta)]m^{-1}(x, \beta)m_\beta(x, \beta),$$

where $m_\beta(x, \beta) = \frac{\partial}{\partial \beta} m(x, \beta)$. Given an i.i.d. sample (x_i, y_i), $i = 1, \ldots, n$, consistent estimators $\widehat{\beta_1}$ and $\widehat{\beta_2}$ of β are found by solving the equations

$$\sum_{i=1}^{n} \psi_j(y_i, x_i, \widehat{\beta_j}) = 0, \quad j = 1, 2,$$

respectively.

The asymptotic covariance matrices of $\widehat{\beta_1}$ and $\widehat{\beta_2}$ are, respectively, given by

$$\Sigma_1 = \left[E\left(v^{-1} m_\beta m'_\beta\right)\right]^{-1},$$

$$\Sigma_2 = \left[E\left(m^{-1} m_\beta m'_\beta\right)\right]^{-1} E\left(v m^{-2} m_\beta m'_\beta\right)\left[E\left(m^{-1} m_\beta m'_\beta\right)\right]^{-1},$$

with v, m, m_β being obvious notational abbreviations.

The problem of comparing Σ_1 with Σ_2 can be reduced algebraically to the situation of the proposition with $\alpha = -1$ and thus to the case mentioned in Remark 2. It is thus seen that

$$\Sigma_1 \leqslant \Sigma_2$$

in the Loewner sense, i.e., $\widehat{\beta_1}$ is asymptotically more efficient than $\widehat{\beta_2}$.

This result also follows from a general theorem by Heyde (1997). However, Shklyar and Schneeweiss [4] can prove, for the Poisson model, the stronger result that $\Sigma_1 < \Sigma_2$ if $\Sigma_\delta \beta_1 \neq 0$.

3.2. *Polynomial model*

Kukush and Schneeweiss [2] and Kukush et al. [3] consider the polynomial model with measurement errors:

$$y = \beta_0 + \beta_1 \xi + \cdots + \beta_k \xi^k + \epsilon,$$

$$x = \xi + \delta,$$

with the same assumptions on ξ and δ (scalar variables this time) as in the previous example and with $\epsilon \sim N(0, \sigma_\epsilon^2)$, ϵ being independent of ξ and δ.

One can again set up an estimating function like ψ_1 with a corresponding estimator of β and an asymptotic covariance matrix, again denoted by $\widehat{\beta_1}$ and Σ_1, respectively.

One can also estimate β via a completely different estimating function:

$$\psi_3(y, x, \beta) = H\beta - h,$$

where h is a vector with components $h_r = y t_r(x)$ and H a matrix with elements $H_{rs} = t_{r+s}(x)$, $r = 0, \ldots, k$, $s = 0, \ldots, k$, and $t_r(x)$ is a polynomial in x of degree r such that $E[t_r(x) \mid \xi] = \xi^r$. The corresponding estimator $\widehat{\beta_3}$ is consistent and has an asymptotic covariance matrix denoted by Σ_3.

It is difficult to compare the relative efficiencies of $\widehat{\beta_1}$ and $\widehat{\beta_3}$. However, in border line cases, when σ_δ^2 and σ_ϵ^2 both become small, a comparison is possible. Let $\sigma_\delta^2/\sigma_\epsilon^2 = \varkappa^2$ and let σ_δ^2 and σ_ϵ^2 go to zero such that $\varkappa > 0$ remains fixed, then,

$$\Sigma_1 = \sigma_\epsilon^2 \left[E\left(v^{-1} \zeta \zeta'\right)\right]^{-1} + \mathrm{O}\left(\sigma_\epsilon^4\right),$$

$$\Sigma_3 = \sigma_\epsilon^2 [E(\zeta\zeta')]^{-1} E(v\zeta\zeta') [E(\zeta\zeta')]^{-1} + \mathrm{O}\left(\sigma_\epsilon^4\right),$$

where $\zeta = (1, \xi, \ldots, \xi^k)'$ and $\upsilon = 1 + \varkappa^2 \left(\frac{\mathrm{d}\zeta'}{\mathrm{d}\xi}\text{ß}\right)^2$. Applying the proposition, one can see that

$$\lim_{\sigma_\epsilon \to 0} \sigma_\epsilon^{-2} \mathrm{tr}\, \Sigma_1 < \lim_{\sigma_\epsilon \to 0} \sigma_\epsilon^{-2} \mathrm{tr}\, \Sigma_2$$

and, by Remark 2, that

$$\lim_{\sigma_\epsilon \to 0} \sigma_\epsilon^{-2} \Sigma_1 \leqslant \lim_{\sigma_\epsilon \to 0} \sigma_\epsilon^{-2} \Sigma_2$$

in the Loewner sense.

Acknowledgements

We thank an anonymous referee for valuable comments on an earlier version of the paper. Support by the Deutsche Forschungsgemeinschaft (SFB 386) is gratefully acknowledged.

References

[1] M. Kendall, A. Stuart, K. Ord, Kendall's Advanced Theory of Statistics, Griffin, London, 1987.
[2] A. Kukush, H. Schneeweiss, A comparison of asymptotic covariance matrices of adjusted least squares and structural least squares in error ridder polynomial regression, Discussion Paper 218 of SFB 386, Department of Statistics, University of Munich, 2000.
[3] A. Kukush, H. Schneeweiss, K. Wolf, Relative efficiency of three estimators in a polynomial regression with measurement errors, J. Statist. Plann. Inference, in press.
[4] S. Shklyar, H. Schneeweiss, A comparison of asymptotic covariance matrices of three estimators in the Poisson regression model with measurement error, Discussion Paper 283 of SFB 386, Department of Statistics, University of Munich, 2002.